Dist: A

REPORT D TATION PAGE

Form Approved
OMB No. 0704-0188

(13)

to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, … ing the collection of information. Send comments regarding this burden estimate or any other aspect of this … to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson … of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

DATE

3. REPORT TYPE AND DATES COVERED
ANNUAL 01 Jun 93 TO 31 May 94

(2)

4. TITLE AND SUBTITLE

DEMODULATION PROCESSES IN AUDITORY PERCEPTION

5. FUNDING NUMBERS

AFOSR-
F49620-93-1-0299

61102F
2313/AS

6. AUTHOR(S)

Dr LawrenceL. Feth

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Speech and Hearing Science
Ohio State University
110 Pressey Hall
Columbus OH 43210

8. PERFORMING ORGANIZATION REPORT NUMBER

AEOSR-TR- 94 0627

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NL
110 Duncan Ave Suite B115
Bolling AFB DC 20332-0001

Dr John F. Tangney

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

A

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The long range goal of this project is the understanding of human auditory processing of information conveyed by complex, time-varying signals such as speech, music or important environmental sounds. Our work is guided by the assumption that human auditory communication is a "modulation - demodulation" process. That is, we assume that sound sources produce a complex stream of sound pressure waves with information encoded as variations (modulations) of the signal amplitude and frequency."picture" and the perception process is modelled as if the listener were analyzing the spectral picture. This approach leads to studies such as "profile analysis" and the power-spectrum model of masking. Our approach leads us to investigate time-varying, complex sounds. We refer to them as dynamic signals and we have developed auditory signal processing models to help ~~guide~~ our experimental work.

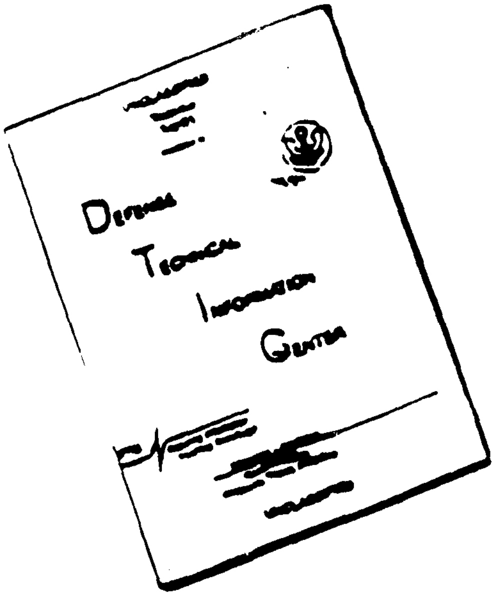
3 9

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)


Defense
Technical
Information
Center

AFOSR--93-1-0299

DEMODULATION PROCESSES IN AUDITORY PERCEPTION

LAWRENCE L. FETH, Ph.D.
SPEECH AND HEARING SCIENCE
OHIO STATE UNIVERSITY
110 PRESSEY HALL
COLUMBUS, OHIO 43210

1 AUGUST 1994

ANNUAL REPORT PERIOD: 1 JUNE 1993 - 31 MAY 1994

PREPARED FOR:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
BOLLING AIR FORCE BASE
WASHINGTON, DC 20332

ANNUAL TECHNICAL REPORT

1 June 1993 - 31 May 1994

Introduction

The long range goal of this project is the understanding of human auditory processing of information conveyed by complex, time-varying signals such as speech, music or important environmental sounds. Our work is guided by the assumption that human auditory communication is a "modulation - demodulation" process. That Is, we assume that sound sources produce a complex stream of sound pressure waves with information encoded as variations (modulations) of the signal amplitude and frequency. The listener's task then is one of demodulation. Much of past psychoacoustics work has been based in what we characterize as "spectrum picture processing." Complex sounds are Fourier analyzed to produce an amplitude-by-frequency "picture" and the perception process is modeled as if the listener were analyzing the spectral picture. This approach leads to studies such as "profile analysis" and the power-spectrum model of masking. Our approach leads us to investigate time-varying, complex sounds. We refer to them as dynamic signals and we have developed auditory signal processing models to help guide our experimental work.

Since the proposal for this project was written in fall 1992, we have re-ordered the sequence of experiments that were proposed. Progress will be described under the headings of the proposal, however, to facilitate relating our work to that document. Also, since the start of the project was June 1 rather than January 1, 1993, some tasks were completed very early in the first year of funding. Since they were not included in the final report of the previous funding period, they have been included in this report.

The TDT equipment purchased in the first year of the project enabled us to generate the complex, time-varying signals in real time. Previously, some signals had to be generated off-line and stored on disk for replay during the experiment. The real time versions of the signals mean that we can run "roving-parameter" paradigms in adaptive tracking procedures. Parameters that can "rove" include signal frequency, duration and amplitude. Roving can be done on a single parameter or in combinations (i.e., roving frequency and amplitude at the same time).

List of research objectives and current progress

A. Single-transition signals - single channel model

1. Frequency Modulated Tones

a. Roving frequency: GLIDE~STEP Discrimination (see paragraph below)

b. Sinusoidal vs. Linear Trajectory

Work began in January 94 on the detection of sinusoidal FM added to a linear FM sweep. To discem the effects of roving frequency on these tasks we incorporated frequency rove into the design of this set of experiments rather than conducting a step vs. glide experiment with roving frequency. Our results have been reported at the June 94 meeting of the Acoustical Society and at the 10th International Symposium on Hearing at Irsee, Bavaria. A manuscript is in progress that will be submitted to the Acoustical Society journal.

c. Slope Discrimination

This area was the topic of the doctoral dissertation of Chien yeh Hsu. His dissertation was completed in the summer quarter 1993, and he reports that a manuscript is in progress. Since July 1993 he has held a post-doctoral position at the University of Illinois.

2. Moving Filter

a. Variation on the GLIDE-STEP Discrimination task

We have by-passed these proposed follow up versions of the original design because we decided that most of the questions raised could be answered by incorporating the variations into sinusoidal plus linear FM designs.

b. Sinusoidal vs. Linear Trajectory

These experiments have not been started. We expect that they will be underway in the second year of the project.

c. Slope Discrimination

The discrimination of the slope of the linear trajectory for a single resonator filter was incorporated into the dissertation of Hsu (see 1.c above). Results were reported at the 1994 meeting of the ARO and we expect a manuscript to be submitted soon.

3. Single Formants from "Real Speech"

This work has not been started. We expect that it will begin at the end of year two or the beginning of year three.

B. Multi-formant signals - multi-channel model

Work on the multi-channel IWAIF model was incorporated into the master's thesis of M. Mokheimer, who applied it to the detection of mixed modulation by human listeners. A presentation based on the thesis is to be given in Cairo in Dec. 1994. Development of the model has continued in year one, with the experimental work on moving filter and "real speech" signals to follow in years two and three.

1. Moving filter multiple-formant signals. (see paragraph above)

2. Multi-formant signals extracted from real speech (see paragraph above)

C. Incorporation of Envelope Cues

We have conducted a series of experiments suggested by the work of Versfeld and Houtsma, and presented preliminary results at the June 94 meeting of the Acoustical Society. As is often the case, the experiments have raised more questions than they answered, and we continue to work on this area.

Participating Professionals

Lawrence L. Feth, Ph.D.	Principal Investigator
Ashok K. Krishnamurthy, Ph.D.	Co-Principal Investigator
Jayanth N. Anantharaman, M.S.	Grad. Research Assoc.
Tao Zhang, M.S.	Grad. Research Assoc.
*Chien yeh Hsu M.S.	Grad. Research Assoc.
*Mohamed Mokheimer	Grad. Research Assoc.

(*no cost to this project period)

Publications and Presentations

Discrimination of broadband, multi-component, common-envelope signals. J. N. Anantharaman, A. K. Krishnamurthy and L. L. Feth, [Abstract: J. Acoust. Soc Amer. Vol. 93, p2387, May 1993].

Short-term IWAIF model for frequency discrimination. A. K. Krishnamurthy and L. L. Feth, [Abstract: J. Acoust. Soc Amer. Vol. 93, p2387, May 1993].

Auditory Discrimination of Frequency Transitions by Human Listeners and a computational Model. Chien-yeh Hsu, Unpublished Doctoral Dissertation, Ohio State University, (August 1993).

Intensity-weighted average of instantaneous frequency as a model for frequency discrimination. J. N. Ananthraraman, A. K. Krishnamurthy and L. L. Feth, J. Acoust. Soc Amer., Vol. 94, 723-729, (1993).

A two-dimensional, independent channels model for complex sound discrimination. T. Zhang, C. Hsu, L. L. Feth and A. K. Krishnamurthy, [Abstract: Seventeenth Midwinter Research Meeting of the Association for Research in Otolaryngology, p 69, Feb. 1994].

An IWAIF model for the detection of mixed modulation. M. Mokheimer, J. N. Anantharaman, L. L. Feth and A. K. Krishnamurthy, Abstract: Seventeenth Midwinter Research Meeting of the Association for Research in Otolaryngology, p71, Feb. 1994].

Discriminability of three-tone, common envelope signals. J. N. Anantharaman, L. L. Feth and A. K. Krishnamurthy, [Abstract: J. Acoust. Soc Amer. Vol. 95, p2964, June 1994].

Detection of frequency modulation in steady and gliding tones. T. Zhang, L. L. Feth and A. K. Krishnamurthy, [Abstract: J. Acoust. Soc Amer. Vol. 95, p2965, June 1994].

Detection of Combinations of Frequency Modulation: An Application of the IWAIF Model. L. L. Feth, A. K. Krishnamurthy and T. Zhang Proceedings of the 10th International Symposium on Hearing, Irsee, Bavaria June 1994 (in press).

A Multi-channel Intensity-Weighted Average of Instantaneous Frequency Model. M. Mokheimer, J. N. Anantharaman, A. K. Krishnamurthy and L. L. Feth, to be presented at the First International Conference on electronics, Circuits and Systems, Cairo, Egypt Dec. 1994. (in press).

Patents and Inventions No patentable inventions have resulted from this work.

Intensity-weighted average of instantaneous frequency as a model for frequency discrimination

Jayanth N. Anantharaman and Ashok K. Krishnamurthy
Department of Electrical Engineering, The Ohio State University, Columbus, Ohio 43210

Lawrence L. Feth
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

(Received 4 March 1992; revised 29 March 1993; accepted 9 April 1993)

The intensity-weighted average of instantaneous frequency (IWAIF) is developed as a model to predict listener performance in tasks primarily requiring frequency discrimination. IWAIF is closely related to the envelope weighted average of instantaneous frequency (EWAIF) model proposed by Feth for similar tasks. The primary difference is that the IWAIF model uses intensity (envelope squared) as the weighting function instead of the envelope. The advantages of IWAIF over EWAIF are that (a) it has a convenient frequency domain interpretation; and (b) it is much simpler to compute than the EWAIF. The IWAIF is the "center of gravity" of the energy spectral density function of the signal.

PACS numbers: 43.66.Ba, 43.66.Fe [HSC]

INTRODUCTION

The envelope-weighted average of instantaneous frequency (EWAIF) model was developed nearly two decades ago by Feth (1974) to account for the discriminability of two-tone complexes. Helmholtz (1954) reported that the pitch of a two-component complex tone is shifted toward the frequency of the component whose amplitude is increased slightly. Helmholtz attributed the pitch shift to fluctuations in the instantaneous frequency of the two-tone complex. Feth and co-workers (Feth, 1974; Feth and O'Malley, 1977; Feth *et al.*, 1982) have studied the discriminability of complementary pairs of two-tone complexes (Voelcker, 1966a,b). Feth showed that the pitch differences are proportional to the EWAIF differences between the complex signals. That is, the EWAIF is calculated for each signal of the pair. The discriminability is predicted as that of pure tones with frequencies at the EWAIF values. The EWAIF model has been used to explain a variety of discrimination tasks where the spectral pitch of the stimulus is the dominant cue. Feth and Stover (1987) extended the model to explain an anomaly in data relating to "profile signals" (Green, 1988). The central theme of this model is that for certain signal pairs, listeners use spectral pitch differences to discriminate between them. Feth's model attempts to quantify the pitch changes observable in the discrimination of complex stimuli. In the case of profile signals, it is assumed that changes in spectral shape of the profile signals produce a noticeable change in the perceived pitch.

This paper introduces a related model, the intensity-weighted average of instantaneous frequency, IWAIF.[1] The main difference between the IWAIF of a signal and its EWAIF is the choice of a weighting function. While the signal envelope is used to weight the instantaneous frequency in the EWAIF calculation, the intensity, which is proportional to envelope squared, is used as the weighting function for the IWAIF calculation. Since both envelop and intensity are non-negative, these weighting function are highly correlated. Thus, similar values are expected fo the EWAIF and IWAIF of the same signal. Indeed, Fet *et al.* (1982) demonstrated that predictions based on envelope and envelope-squared weights, as well as, rms versus arithmetic averaging made little difference in th weighted-frequency average calculations.

The advantage of the IWAIF lies in computationa efficiency and accuracy. Analytical calculation of th EWAIF requires a bit of algebra and trigonometry to derive expressions for the envelope and the instantaneou frequency. For two components (Feth, 1974) the analytical solution is straightforward; for three components Kid *et al.* (1991) have produced an analytic solution. For mor than three components, the analytic approach is dauntin

Discrete approximations to the EWAIF calculatio present a new set of problems. The instantaneous fr quency must be calculated by taking the derivative of i stantaneous phase, a highly noise-sensitive process. Als the EWAIF may require the division of two near-zer quantities, which may lead to underflow errors in finit word length representations of the values. The IWAIF formulation can be transformed into the frequency domai (Anantharaman *et al.*, 1991). In addition to avoiding t calculation problems of the time-domain version ‹ EWAIF, the IWAIF provides both computational efficiency and a novel interpretation of its value. The IWAI is equivalent to the spectral "center of gravity."

First, the time and frequency domain representatio of the EWAIF are presented. The IWAIF of a signal then defined, and its representation in the frequency d main is derived. The performance of the IWAIF model then compared to that of the EWAIF model in a numb of psychoacoustic tasks.

I. EWAIF MODEL

A. EWAIF in the time domain

In general, a finite energy real signal $s(t)$ which has a Fourier transform

$$S(f)=\mathscr{F}[s(t)]=\int_{-\infty}^{\infty} s(t)e^{-2\pi jft}\,dt \tag{1}$$

can be represented as (McGillem, 1979; Voelcker, 1966a,b),

$$s(t)=e(t)\cos\phi(t), \quad 0<t<T, \tag{2}$$

$$=\mathrm{Re}[e(t)e^{j\phi(t)}], \tag{3}$$

where $e(t)$ is the instantaneous envelope, $\phi(t)$ is the instantaneous phase, and Re[ ] denotes the real part operator. The instantaneous frequency, $f(t)$ is defined as

$$f(t)=\frac{1}{2\pi}\frac{d\phi(t)}{dt}. \tag{4}$$

Such a representation of $s(t)$ is not unique. For example, $e(t)$ can be chosen to satisfy (3) for an arbitrary $\phi(t)$. A unique $e(t)$ and $\phi(t)$ can be assured by imposing an additional constraint, namely, that the real and imaginary parts of the complex signal $e(t)e^{j\phi(t)}$ form a Hilbert transform pair. Such a complex signal is termed analytic and has certain useful properties. Thus, the analytic signal corresponding to the real signal $s(t)$ can be written as

$$m(t)=s(t)+j\hat{s}(t), \tag{5}$$

$$=|m(t)|e^{j\phi(t)}, \tag{6}$$

where

$$\hat{s}(t)=\mathscr{H}[s(t)] \tag{7}$$

is the Hilbert transform of $s(t)$.

The envelope and instantaneous frequency functions, $e(t)$ and $f(t)$, can be defined in terms of $s(t)$ and $\hat{s}(t)$ as

$$e(t)=|m(t)|=[s^2(t)+\hat{s}^2(t)]^{1/2}, \tag{8}$$

$$\phi(t)=\arctan\left(\frac{\hat{s}(t)}{s(t)}\right), \tag{9}$$

$$f(t)=\frac{1}{2\pi}\frac{s(t)\hat{s}'(t)-s'(t)\hat{s}(t)}{s^2(t)+\hat{s}^2(t)}. \tag{10}$$

The envelope-weighted average of instantaneous frequency (EWAIF) of $s(t)$ is defined as

$$\mathrm{EWAIF}[s(t)]=\frac{\int_0^T e(t)f(t)dt}{\int_0^T e(t)dt}. \tag{11}$$

A common method of calculating the EWAIF of a signal is to determine the envelope and instantaneous frequency functions using (8), (10), and computing the required integrals in (11). However, there are some computational problems when we adopt this method for calculating the EWAIF of broadband signals. Note that the expression (10) for $f(t)$ involves differentiation which is a highly noise-sensitive operation.

B. Frequency domain representation of EWAIF

Alternatively $f(t)$ can be expressed in terms of the analytic signal $m(t)$, alone by rewriting (6) as

$$\ln m(t)=\ln|m(t)|+j\phi(t). \tag{12}$$

Hence,

$$\phi(t)=\mathrm{Im}[\ln m(t)], \tag{13}$$

where Im denotes the imaginary part operator,

$$f(t)=\frac{1}{2\pi}\mathrm{Im}\left(\frac{m'(t)}{m(t)}\right). \tag{14}$$

Inserting the above equations in the expression for EWAIF (11) we have

$$\mathrm{EWAIF}[s(t)]=\frac{1}{2\pi}\frac{\int_0^T|m(t)|\mathrm{Im}[m'(t)/m(t)]dt}{\int_0^T|m(t)|dt}. \tag{15}$$

This can be expressed in terms of the Fourier transform of $\sqrt{m(t)}$ as (see Appendix A for the derivation)

$$\mathrm{EWAIF}[s(t)]=2\frac{\int_{-\infty}^{\infty}f|M_S(f)|^2df}{\int_{-\infty}^{\infty}|M_S(f)|^2df}, \tag{16}$$

where $M_S(f)=\mathscr{F}[\sqrt{m(t)}]$.

The EWAIF is thus the frequency of the "center of gravity" of $|M_S(f)|^2$. While this is an interesting observation, it is of little use in the computation of the EWAIF of a signal. Indeed, in order to obtain $M_S(f)$, the square root of a complex signal has to be computed. In computing $\sqrt{m(t)}$, we have to be careful to choose the principal branch of the square root. This is similar to the phase unwrapping problem encountered in signal processing. Further, because $f(t)$ has an $e(t)$ term in the denominator, care must be taken in computing the instantaneous frequency at points where the envelope is zero or near zero. This involves computing a limit of the ratio of two functions which approach zero rather than a simple division.

II. IWAIF

In computing the EWAIF, the envelope of the signal is used as the weighting function for finding the average of the instantaneous frequency. Other weighting functions may model listener discriminability as well. Indeed, predictions based on an envelope-squared (intensity) weighted model have performed as well as an envelope weighted model (Feth *et al.*, 1982). This is to be expected as intensity is the square of the envelope, a non-negative function. To this end, let us investigate the intensity-weighted (arithmetic) average of instantaneous frequency (IWAIF) of a signal. The IWAIF of $s(t)$ is defined as

$$\mathrm{IWAIF}[s(t)]=\frac{\int_0^T e^2(t)f(t)dt}{\int_0^T e^2(t)dt}, \tag{17}$$

where $e(t)$ and $f(t)$ are as defined in Eqs. (8) and (10), respectively.

The above definition of IWAIF was motivated by our previous work with EWAIF. Equation (17) can also be arrived at in an alternative way. Suppose that a suitable frequency f_0 is to be found such that $s(t)$ represents a modulated wave of the form

$$s(t)=e(t)\cos[2\pi f_0 t+\theta(t)], \tag{18}$$

$$=\mathrm{Re}[m(t)]. \tag{19}$$

$e(t)$ is thought of as the envelope of $s(t)$ and $\theta(t)$ as its phase. $m(t)$ is the complex analytic signal corresponding to $s(t)$ as given in Eq. (5). For narrow band $e(t)$ and $\theta(t)$ this represents the modulation of a sinusoidal carrier wave of frequency f_0. The instantaneous frequency of the signal is

$$f(t)=f_0+\frac{1}{2\pi}\frac{d\theta(t)}{dt}. \tag{20}$$

The choice of f_0 can be arbitrary so long as the mathematical relations remain valid. The most common choice (McGillem, 1979) is to select f_0 such that it is the center of gravity of $|M(f)|^2$. This corresponds to the center of gravity of the positive frequency portion of the energy density spectrum of the signal. The required value for f_0 is that value which minimizes the following integral:

$$\int_0^\infty (f-f_0)^2|M(f)|^2df, \tag{21}$$

which is the same as the IWAIF of the signal $s(t)$.

A. Frequency domain representation of IWAIF

Much of the discussion in this section follows that in Anantharaman (1992). We can rewrite (17) in terms of $m(t)$ as

$$\mathrm{IWAIF}[s(t)]=\frac{1}{2\pi}\frac{\int_0^T|m(t)|^2\,\mathrm{Im}[m'(t)/m(t)]dt}{\int_0^T|m(t)|^2dt}. \tag{22}$$

Invoking Parseval's relation this becomes (see Appendix B)

$$\mathrm{IWAIF}[s(t)]=\frac{\int_{-\infty}^{\infty}f|M(f)|^2df}{\int_{-\infty}^{\infty}|M(f)|^2df}. \tag{23}$$

This can be further simplified by taking advantage of the one-sided nature of $M(f)$ and its relation to $S(f)$. Equation (23) then becomes

$$\mathrm{IWAIF}[s(t)]=\frac{\int_0^\infty f|S(f)|^2df}{\int_0^\infty|S(f)|^2df}. \tag{24}$$

Thus, the IWAIF of a real signal is located exactly at the "center of gravity" of the positive portion of its energy density spectrum. The above frequency domain representation (24) provides a simple and efficient procedure for computing the IWAIF of a signal. Compare this with (16) which is the frequency domain expression for the EWAIF. Using (24) eliminates most of the difficulties encountered in computing the EWAIF of the signal. The IWAIF is completely described by the energy spectrum of the signal alone. This obviates the need to compute a Hilbert transform and a derivative. All that needs to be computed is the Fourier transform of $s(t)$. This can be done efficiently using the FFT algorithm.

Suppose $s(t)$ is sampled at a rate F_s to yield N samples, $s[n]$, $n=0,1,\ldots,N-1$, and its N-point FFT is $S[k]$, $k=0,1,\ldots,N-1$. Then, the IWAIF of $s(t)$ can be computed as

$$\mathrm{IWAIF}[s(t)]\approx\frac{\Sigma_{k=0}^{(N/2)-1}k\Delta f|S(k)|^2\Delta f}{\Sigma_{k=0}^{(N/2)-1}|S(k)|^2\Delta f}$$

$$=\Delta f\frac{\Sigma_{k=0}^{(N/2)-1}k|S(k)|^2}{\Sigma_{k=0}^{(N/2)-1}|S(k)|^2}, \tag{25}$$

where $\Delta f=F_s/N$ is the frequency spacing between samples of the FFT.

III. COMPARISON OF IWAIF AND EWAIF PREDICTIONS WITH PSYCHOACOUSTIC RESULTS

As mentioned earlier, the main difference between the IWAIF of a signal and its EWAIF is in the choice of a weighting function. While the envelope is used to weight the instantaneous frequency in calculating the EWAIF, the intensity (envelope squared) is used as the weighting function in IWAIF calculations. Since both envelope and intensity are non-negative and the latter is the square of the former, the weighting functions are highly correlated. Thus, similar values are expected for the EWAIF and IWAIF of a signal. For a simple sinusoid, both the EWAIF and the IWAIF values are equal to the tone frequency f_0. For a combination of two tones of the same amplitude the EWAIF and IWAIF values are again equal and are located at the mean of the two frequencies. It is difficult to calculate the EWAIF of a combination of three or more tones analytically. However, the IWAIF of an N-component complex can be easily calculated. Assuming T to be much larger than the maximum of all the tone periods, the IWAIF of a sum of sinusoids such as

$$s(t)=\sum_i a_i\cos(2\pi f_i t), \quad 0<t<T \tag{26}$$

is approximately equal to the weighted mean

$$\mathrm{IWAIF}[s(t)]=\frac{\Sigma_i a_i^2 f_i}{\Sigma_i a_i^2}. \tag{27}$$

The above relation would have been exact had the sinusoids extended in time from $-\infty$ to $+\infty$. For finite duration signals being considered here the approximation gets better as the duration T increases.

To illustrate the comparable predictions of EWAIF and IWAIF models, we present model predictions for normal-hearing listener performance in frequency discrimination and spectral pitch matching experiments reported previously.

A. Two-component, common envelope complex tones

Feth *et al.* (1982) asked four well-practiced listeners to distinguish between pairs of common envelope, complex tones (see Voelcker, 1966a,b for a discussion of common

TABLE I. EWAIF and IWAIF values for complementary Voelcker signal pairs differing in frequency by Δf Hz and in intensity by ΔI dB. The average $P(C)$ values for four subjects at a center frequency of 2000 Hz are also shown. Predicted differences are independent of the center frequency. Note that an increase in $P(C)$ results in a corresponding increase in ΔEWAIF or ΔIWAIF which is just as expected from the frequency discrimination data for pure tones.

			Predicted pitch match differences	
Δf (Hz)	ΔI (dB)	Average $P(C)$	ΔEWAIF (Hz)	ΔIWAIF (Hz)
10	0.5	63.5%	0.8	0.6
	1.0	69.0%	1.4	1.2
	3.0	83.0%	3.7	3.3
20	0.5	[illegible] 5%	1.6	1.2
	1.0	71.0%	2.9	2.3
	3.0	83.0%	7.3	6.7
50	0.5	63.8%	4.1	2.9
	1.0	77.3%	7.2	5.8
	3.0	87.3%	18.2	16.7
100	0.5	67.8%	8.1	5.8
	1.0	77.8%	14.3	11.5
	3.0	96.0%	36.5	33.2

envelope signals). In addition, the listeners were required to match the pitch they heard for each signal in the pair to that in a two-component signal with equal amplitude components. The equal-amplitude signal was adjustable along the frequency axis to enable the spectral pitch match. In that study, Feth *et al.*, presented both predicted pitch values, based on EWAIF calculations and the averaged pitch match for each signal. They assumed that the EWAIF for each of the signals of a given pair represented the frequency of a sinusoid that would produce the same spectral pitch. The difference between EWAIF values was assigned a predicted percentage of correct discriminations. Feth (1974) had found that such predictions were good indicators of listener performance for discrimination between pairs of common envelope signals. In Table I, we present selected discrimination results from Feth *et al.* (1982) along with predicted EWAIF and IWAIF model predictions. As previously noted by Feth *et al.*, differences in IWAIF values for a given signal pair tend to be slightly smaller than the equivalent EWAIF values, but the differences are considered negligible.

B. Simultaneous amplitude and frequency modulation

Iwamiya *et al.* (1984) studied the principal pitch heard by listeners in signals designed to approximate vibrato in musical sounds. Note that principal pitch is another term for spectral pitch. Complex sounds were generated by modulating a sinusoidal carrier both in frequency and amplitude with the same modulating signal. The carrier (at frequencies of 440, 880, or 1500 Hz) was modulated by a low-frequency (6 Hz) triangular waveform. Thus, if $D_{\rm AM}$ is the "degree of AM"[2] and $E_{\rm FM}$ is the "extent of FM in cents,"[3] the modulated signal for a carrier f_c is given by

$$s(t)=[1+D_{\rm AM}m(t)]\times\cos\left(2\pi f_c t+0.5E_{\rm FM}\int_0^t m(\tau)d\tau\right), \qquad (28)$$

where $m(t)$ is the modulating signal.

Listeners were asked to match the pitch they heard in these modulated tones to that of a pure tone. The experiment was conducted for AM and FM modulations presented "in-phase" and "out-of-phase." That is, the modulations were "in-phase" when a frequency increase was accompanied by an amplitude increase. Out-of-phase, then, meant that frequency increase accompanied an amplitude decrease. Also, they conducted some pitch matches with the degree of AM set to 1.0 with the extent of FM taking values of 0, 25, 50, and 100 cents. Other trials held the extent of FM at 100 cents while the degree of AM was 0.0, 0.50, 0.75, or 1.00.

Results taken from Iwamiya *et al.*, are plotted in Fig. 1. Also shown in Fig. 1 are spectral pitch values predicted by the EWAIF and IWAIF models. Note that the listeners exhibit a small negative bias. Matches to simple AM tones, which should be at the carrier frequency (i.e., zero difference from f_c) fall a few cents below. The EWAIF and IWAIF models cannot account for this bias; however, they both produce predicted spectral pitch match results for "in-phase" and "out-of-phase" conditions that agree with listener performance as extent of modulation is increased. In general, the models predict somewhat larger pitch differences between signal pairs (in-phase versus out-of-phase) as compared to the listeners.

C. Application to signals used in profile analysis studies

Feth and Stover (1987) attempted to extend the EWAIF model to the complex signals used in many of the early profile analysis experiments. In those studies the listener was asked to determine which of two signals, consisting of a number of sinusoids, contained a small increment to the amplitude of the sinusoid in the center. To deter the use of absolute intensity discrimination in determining which complex signal contains the increment, the overall level of each presentation is selected at random from a range of level values. This is commonly called a "roving level" paradigm. The assumption is that since the listeners will be unable to use simple (absolute) intensity cues, they will be forced to base discrimination decisions on the difference in overall spectral shape, or profile, of the complex signals. This assumption ignores the possibility that listeners may respond to other cues available in the incremented signal. For example, interactions among the inharmonically spaced sinusoids that make up the complex signal can lead to frequency modulations (FM). Further, adding a small increment to the amplitude of one sinusoid in the complex can lead to a change in the FM produced by the interaction. Such changes in FM may be audible as subtle pitch shifts. The size of the frequency shift produced by an increment to one component depends only on the

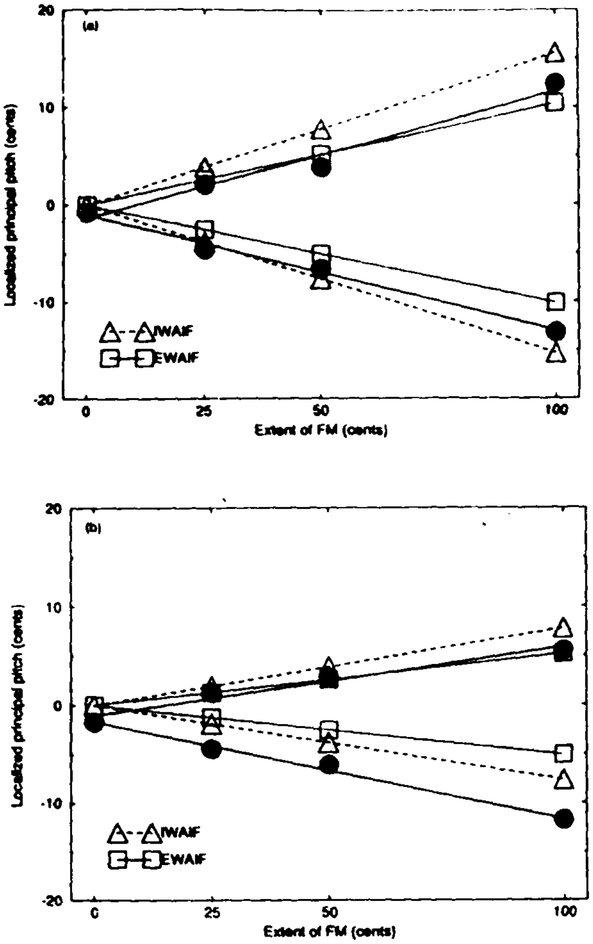


FIG. 1. Localized principal pitch, EWAIF and IWAIF of FM-AM tones as a function of the extent of FM with the frequency and amplitude modulations both in-phase (lines with positive slope) and anti-phase (lines with negative slope) (Iwamiya *et al.*, 1984). The solid lines are a regression line fit to the collected data represented by solid c·rcles. The open triangles and open squares represent the IWAIF and the EWAIF predictions, respectively. (a) 440-Hz carrier frequency, (b) 880-Hz carrier frequency.

relative amplitudes of the components, not the absolute levels. Thus, the FM produced would be unaffected by "roving level" procedures.

Early profile investigators were puzzled by an anomalous result that occurred when discrimination performance was observed as the number of components in the profile signal was increased (Green *et al.*, 1983; Green *et al.*, 1984). As the number of components was increased from 3 to 11 sinusoids, the just-detectable increment was found to be progressively smaller. That is, it appeared that listeners were more sensitive to an increment in 1 of 11 components than they were to an increment in 1 of 3 or 5 components. The puzzle was to explain how adding additional components to the complex signal could lead to such enhanced performance. However, if we consider the FM produced by an increment to the center component in the complex signal, we may find an alternative explanation for listener behavior.

TABLE II. EWAIF and IWAIF based predictions of pitch differences between multicomponent equal-amplitude profile signals and their incremented counterparts (with threshold increments). Note that while threshold increment gets smaller as more components are added, there is little change in the ΔEWAIF and ΔIWAIF values indicating equal listener performance.

Number of components in profile complex	Threshold increment ΔI (dB)	Model predictions ΔEWAIF (Hz)	ΔIWAIF (Hz)
3	−0.1	21.1	17.3
5	−4.2	29.5	25.8
7	−11.2	21.8	18.0
9	−13.0	23.7	20.0
11	−13.9	33.6	29.8

The increment in the amplitude of the signal component leads to a small difference in the EWAIF values calculated for the standard and the target complex sounds. These EWAIF differences reflect the difference in FM between these sounds. Feth and Stover (1987) showed that the EWAIF difference between the just-detectable target and the standard was approximately the same, independent of the number of components in the profile signals. Thus, while a profile analysis approach is unable to explain the listeners' improvement in performance with increasing numbers of components, the EWAIF provides a successful explanation. Progressively smaller increments in the amplitude of the central sinusoid of complex signals made up of larger numbers ... components, leads to approximately the same difference in EWAIF. Feth and Stover (1987) argued that the detection of frequency modulation was a more parsimonious explanation of the phenomenon. Table II gives the comparison of just-detectable amplitude increments across component number with the corresponding EWAIF and IWAIF values that each increment produces. Note that EWAIF and IWAIF predictions are quite comparable.

IV. CONCLUSIONS

The intensity-weighted average of instantaneous frequency (IWAIF) model has been presented as an alternative to the envelope-weighted average of instantaneous frequency (EWAIF) model. Calculation of the EWAIF of a signal involves determining the envelope and the instantaneous frequency functions of the signal separately. This can be computationally cumbersome especially as the bandwidth of the signal gets wider. The IWAIF of a signal, on the other hand, can be expressed solely in terms of the magnitude spectrum of the signal. Such a frequency domain representation provides a fast and efficient method to compute the IWAIF of a signal using the FFT algorithm.

The IWAIF model was tested on three sets of stimuli viz. Voelcker's complementary two-tone complexes used in experiments by Feth and co-workers (Feth, 1974; Feth *et al.*, 1982; Feth and O'Malley, 1977), FM-AM tones used by Iwamiya *et al.* (1984), and profile signals used by Green *et al.* (1984). The performance of the IWAIF model was found to be comparable to that of the EWAIF model.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. 89-0227 from the Air Force Office of Scientific Research and by a grant from the Ohio Regents Research Challenge Award.

APPENDIX A: EWAIF

The frequency domain representation of the EWAIF can be derived as follows. The EWAIF of $s(t)$ can be written as

$$\text{EWAIF}=\frac{\int_0^T e(t)f(t)dt}{\int_0^T e(t)dt}$$

$$=\frac{1}{2\pi}\frac{\int_0^T|m(t)|\text{Im}[m'(t)/m(t)]dt}{\int_0^T|m(t)|dt}. \tag{A1}$$

Consider the numerator,

$$\int_0^T e(t)f(t)dt=\frac{1}{2\pi}\int_0^T|m(t)|\text{Im}\left(\frac{m'(t)}{m(t)}\right)dt, \tag{A2}$$

$$=\frac{1}{2\pi}\text{Im}\int_0^T\sqrt{m(t)m^*(t)}\,\frac{m'(t)}{m(t)}dt, \tag{A3}$$

where $m^*(t)$ denotes the complex conjugate of $m(t)$

$$=\frac{1}{2\pi}\text{Im}\int_0^T[\sqrt{m(t)}]^*\frac{m'(t)}{\sqrt{m(t)}}dt, \tag{A4}$$

$$=\frac{1}{\pi}\text{Im}\int_0^T[\sqrt{m(t)}]'[\sqrt{m(t)}]^*\,dt. \tag{A5}$$

Applying the theorem for the Fourier transform of the derivative of a signal and invoking Parseval's theorem, the numerator can be expressed as

$$\int_0^T e(t)f(t)dt=\frac{1}{\pi}\text{Im}\int_{-\infty}^{\infty}j2\pi f M_S(f)M_S^*(f)df, \tag{A6}$$

$$=2\int_{-\infty}^{\infty}f|M_S(f)|^2df, \tag{A7}$$

where $M_S(f)=\mathscr{F}[\sqrt{m(t)}]$ is the Fourier transform.

Similarly, the denominator can be expressed as

$$\int_0^T|m(t)|dt=\int_0^T\sqrt{m(t)}[\sqrt{m(t)}]^*dt, \tag{A8}$$

$$=\int_{-\infty}^{\infty}|M_S(f)|^2df. \tag{A9}$$

Hence,

$$\text{EWAIF}=2\,\frac{\int_{-\infty}^{\infty}f|M_S(f)|^2df}{\int_{-\infty}^{\infty}|M_S(f)|^2df}. \tag{A10}$$

APPENDIX B: IWAIF

In order to derive the frequency domain representation of IWAIF consider

$$\text{IWAIF}[s(t)]=\frac{\int_0^T e^2(t)f(t)dt}{\int_0^T e^2(t)dt}$$

$$=\frac{1}{2\pi}\frac{\int_0^T|m(t)|^2\,\text{Im}[m'(t)/m(t)]dt}{\int_0^T|m(t)|^2dt}. \tag{B1}$$

In the frequency domain the numerator can be expressed as

$$\int_0^T e^2(t)f(t)dt=\frac{1}{2\pi}\text{Im}\int_0^T m(t)m^*(t)\left(\frac{m'(t)}{m(t)}\right)dt, \tag{B2}$$

$$=\frac{1}{2\pi}\text{Im}\int_0^T m'(t)m^*(t)dt, \tag{B3}$$

$$=\frac{1}{2\pi}\text{Im}\int_{-\infty}^{\infty}j2\pi f M(f)M^*(f)df, \tag{B4}$$

$$=\int_{-\infty}^{\infty}f|M(f)|^2df. \tag{B5}$$

The expression for the Fourier transform of the differential of a signal as well as Parseval's relation were made use of in the foregoing simplification. Again, by Parseval's relation, the denominator is

$$\int_0^T e^2(t)dt=\int_0^T|m(t)|^2dt, \tag{B6}$$

$$=\int_{-\infty}^{\infty}|M(f)|^2df. \tag{B7}$$

Hence,

$$\text{IWAIF}[s(t)]=\frac{\int_{-\infty}^{\infty}f|M(f)|^2df}{\int_{-\infty}^{\infty}|M(f)|^2df}. \tag{B8}$$

[1] The IWAIF has also been referred to as the squared-envelope-weighted average of instantaneous frequency (SEWAIF).

[2] Degree of amplitude modulation (AM) is given by $(A_{\max}-A_{\min})/(A_{\max}+A_{\min})$.

[3] Extent of frequency modulation (FM) is given in cents as $1200\log_2(f_{\max}/f_{\min})$.

Anantharaman, J. N. (1992). "A multichannel signal processing model for complex sound discrimination," Master's thesis, The Ohio State University, Columbus, OH.

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (1991). "Auditory processing of complex signals using the multichannel EWAIF," J. Acoust. Soc. Am. 89, 1938–1939 (A).

Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," Percept. Psychophys. 15, 375–378.

Feth, L. L., and O'Malley, H. (1977). "Two-tone auditory spectral resolution," J. Acoust. Soc. Am. 62, 940–947.

Feth, L. L., O'Malley, H., and Ramsey, J., Jr. (1982). "Pitch of unresolved, two-component complex tones," J. Acoust. Soc. Am. 72, 1403–1412.

Feth, L. L., and Stover, L. J. (1987). "Demodulation processes in auditory perception," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 76–86.

Green, D. M. (1988). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination," J. Acoust. Soc. Am. 73, 639–643.

Green, D. M., Mason, C. R., and Kidd, G., Jr. (1984). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. 75, 1163–1167.

Iwamiya, S., Nishikawa, S., and Kitamura, O. (1984). "Perceived principal pitch of FM-AM tones when the phase difference between frequency modulation and amplitude modulation is in-phase and anti-phase," J. Acoust. Soc. Jpn. 5, 59–69.

Kidd, G., Jr., Mason, C. R., Uchanski, R. M., Brantlet, M. A., and Shah, P. (1991). "Evaluation of simple models of auditory profile analysis using random reference spectra," J. Acoust. Soc. Am. 90, 1340–1354.

McGillem, C. D. (1979). Hilbert Transforms and Analytic Signals (unpublished notes on signal processing, Purdue University).

Voelcker, H. B. (1966a). "Toward a unified theory of modulation—Part I: Phase-envelope relationships," Proc. IEEE 54, 340–353.

Voelcker, H. B. (1966b). "Toward a unified theory of modulation—Part II: Zero manipulation," Proc. IEEE 54, 735–755.

von Helmholtz, H. L. F. (1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York), 2nd English edition, p. 165 and Appendix XIV.

10th International Symposium on Hearing
Irsee, Germany

Detection of Combinations of Frequency Modulation: An Application of the IWAIF Model

Lawrence L. Feth, Ashok K. Krishnamurthy* and Tao Zhang

*Speech and Hearing Science and *Electrical Engineering, The Ohio State University*
Columbus, Ohio 43210 USA

1 Introduction

Our work has been guided by the underlying assumption that human auditory communication is a modulation - demodulation process. That is, we assume that sources produce a complex stream of sound pressure waves with information encoded as variations (i.e., modulation) of signal amplitude and frequency. Speech, music and most environmentally important sounds can be described in this way. Recently, Maragos, et al., (1992) have shown that an energy-tracking operator can be applied to speech signals to produce an algorithm that tracks speech formant amplitude and frequency. The result of the energy-tracking operation is an amplitude-by-frequency product that is quite similar to the EWAIF—IWAIF calculations used in our previous work (Anantharaman, et al., 1993). Earlier work by Teager and Teager (1990) showed that the production of speech could be modeled as the modulation of amplitude and frequency of each formant. Fineberg, et al., (1992) have also begun to apply the modulation model to speech recognition problems.

The human listener's task is then modeled as one of demodulating the sound stream. Much of the past work in psychoacoustics might be characterized as "spectrum picture processing." That is, complex sounds are Fourier-analyzed into an amplitude-by-frequency picture. The experimenter then models auditory perception as a process of analyzing this "spectrum picture." The work on "profile analysis" (see Green, 1988) could be described in this manner. The spectrum picture processing approach leads to studies of broad bandwidth, complex sounds in masking or discrimination experiments. Our "mo-dem" approach leads us to investigate time-varying, complex sounds. We suggest that understanding the auditory processing of such signals with dynamic spectra is essential to better understanding human auditory perception.

1.1 Previous work with STEP-GLIDE signals

The Envelope-Weighted Average of Instantaneous Frequency (EWAIF) model was originally developed to predict the spectral pitch of narrow bandwidth signals for normal hearing human listeners (Feth, 1974; Feth, et al., 1982). We moved to the study of frequency transitions (Feth, et al., 1989) because formant transitions are essential in the description of consonants in speech (e.g., Liberman, et al., 1956). Our experimental work was first designed to determine the appropriate width of the ear's temporal window for the processing of such signals. Several determinations of the limits of auditory temporal acuity are given in the literature (e.g., Green, 1991, 1973 a & b, 1985). Moore and his colleagues have reported the equivalent rectangular duration (ERD) for a temporal window that is the time-domain analog of the roex filter shape of the peripheral filter bank (Moore, et al., 1988; Plack and Moore, 1990, 1991). Little of this previous work has used dynamic signals, that is, signals with frequency transitions. To determine a measure of temporal acuity for such signals, we devised a discrimination task in which listeners were asked to distinguish between a tone frequency modulated over a linear trajectory (a GLIDE) and one covering the same frequency change via a multiple-step trajectory (the STEP). Details are given in Madden and Feth (1992), but the essential result was that the just-discriminable step was approximately 7 to 10 ms for frequencies of 2 kHz and below. Above 2 kHz, the just-discriminable step becomes longer.

Application of the original EWAIF model to dynamic signals proved to be difficult because of the complexity of the EWAIF calculation. The result was the IWAIF model (Anantharaman, et al., 1993). Here, intensity is used to weight the frequency values. The advantage is that the IWAIF can be calculated in the frequency domain using an FFT, whereas the EWAIF was calculated in the time domain. In addition to the great improvement in computational efficiency, the IWAIF form of the model has led to an interpretation of the model output that appears to have wide applicability in our further understanding of auditory signal processing. The result of the IWAIF calculation is the "center of gravity" of the signal spectrum. Such a simple concept has great intuitive appeal for predicting the locus of spectral pitch for many sounds. It is also the basis for the "perceptual formant" suggested by Chistovitch (1979) as the determinant of vowel quality.

A short-term model (ST-IWAIF) can be applied to signals with frequency transitions (Krishnamurthy and Feth, 1993). For example, the STEP and GLIDE signals have nearly the same long-term spectrum, and consequently, the same long-term IWAIF values. Nonetheless, they are easily discriminable, indicating that human listeners are able to utilize cues that are more short-term in nature. To explain such data, it is necessary to introduce the ST-IWAIF model. The model is based on the assumption that the listener can track the changing IWAIF of a dynamic signal and use it as a potential cue for discriminating between two signals. The ST-IWAIF of a signal at time t_0 is determined by the spectral properties of the signal in a small time window of duration T_w around t_0. Let $s_1(t)$ and $s_2(t)$ be the two signals, both of duration T that are to be discriminated. The listener is assumed to track the ST-IWAIF values $I_1(t)$and $I_2(t)$, $0 \leq t \leq T$, of $s_1(t)$ and $s_2(t)$, respectively. Further, we assume that there is internal noise in the auditory system that limits our ability to track frequency. This internal noise is modeled as an additive noise $w(t)$ that corrupts the true ST-IWAIF values. We assume that $w(t)$ is white, zero-mean, Gaussian noise with a power spectral density of $\frac{N_0}{2}$. Since the IWAIF is essentially a signal frequency parameter, we suggest that pure tone frequency-difference limen data be used for that purpose. Thus, if the frequency DL for a tone of duration T at the IWAIF frequency is Δ, we suggest that

$$\frac{N_0}{2} = \Delta^2. \tag{1}$$

Given the above assumptions, the d' for this model is given by

$$d' = \sqrt{\frac{1}{\Delta^2}\int_0^T (I_2(t) - I_1(t))^2 dt}. \qquad (2)$$

The ST-IWAIF model was applied to the results of the original STEP-GLIDE discrimination task with good results. Figure 1 shows the duration of an individual step for STEP signals that were discriminable from GLIDE signals on 75% of the trials. Open symbols represent results averaged for four listeners over signal duration ranging from 25 to 100 ms. Performance is collapsed over signal duration and overall frequency excursion, to be characterized by the rate of transition. Transition rates were 2-, 4- and 8-Hz/ms. Filled symbols represent the performance predicted by the ST-IWAIF model. Model parameters were adjusted so that predictions and data were matched at 1 kHz for the 8 Hz/ms conditions. Those same parameters were then used to predict performance for all other conditions. In general, the model predicts average listener performance very well. At 250 Hz it predicts better performance (i.e., shorter step size). At 4 kHz, the model predicts poorer performance (larger step size) than our listeners obtained.

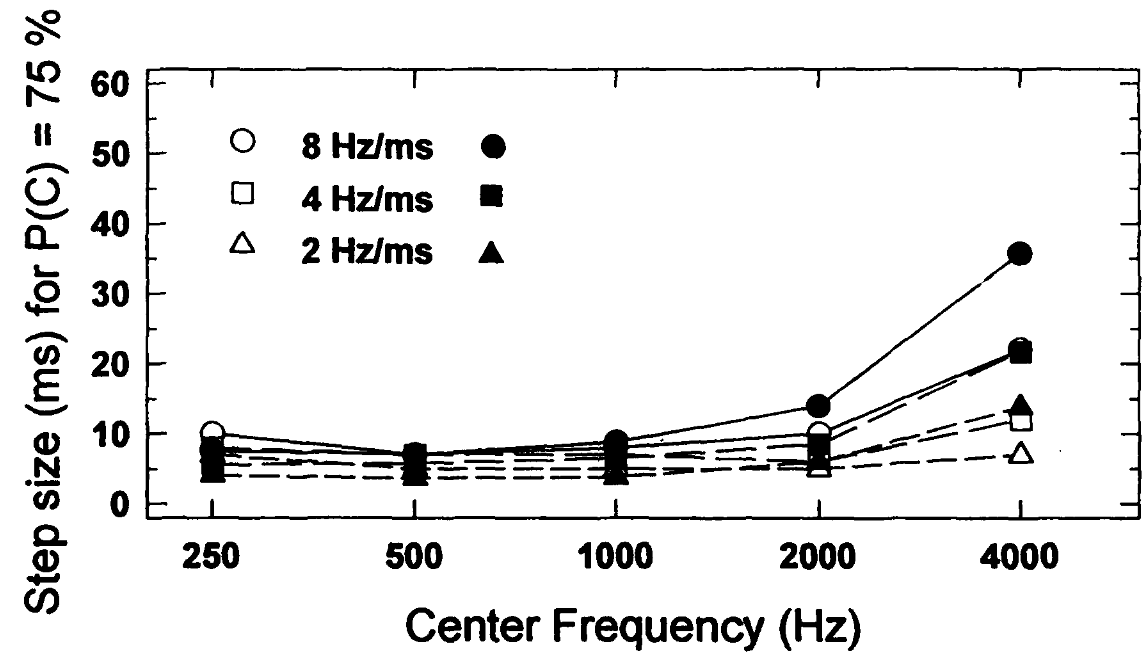


Figure 1. Just-discriminable step size in ms. Open symbols are averaged for four listeners with normal hearing. Circles are for a sweep rate of 8 Hz/ms, squares are 4 Hz/ms and triangles are 2 Hz/ms. Filled symbols are the predicted values obtained from the ST-IWAIF model.

1.2 Madden's model for FM glide discrimination

Recently, Madden (1994) extended the investigation of temporal processing of FM glides. Using an adaptive 2AFC task, he determined the smallest frequency increase between successive steps at which the STEP signal could just be distinguished from the GLIDE. Madden modeled his results using an intensity-based model: A bank of bandpass filters each followed by a non-linearity, temporal window and level detector. Similar models have been used by several investigators of auditory temporal acuity (e.g., Viemeister, 1979; Forrest and Green, 1987; Shailer and Moore, 1987; Green and Forrest, 1988). Madden's modeling indicated that the equivalent rectangular duration (ERD) of the temporal window was about 5 ms for signals ranging from 250 Hz to 6 kHz. However, he was forced to allow detector efficiency to vary substantially over the frequency range to obtain a fit to his data. This is in marked contrast to the detector efficiency reported for temporal acuity of signals without FM.

1.3 IWAIF model predictions of Madden's results

The ST-IWAIF model was applied to Madden's results. The model predictions are shown in Figure 2 along with the averaged data from Madden's paper. The ST-IWAIF model predicts slightly better performance than Madden's listeners obtained. Given that the listeners may not be 100% efficient, the prediction of slightly better performance is not unexpected. Behavior of the model predictions is in line with expectations, except for signals with a large number of steps. For the difficult discrimination of nine, ten or eleven steps versus the seventeen steps in the standard signal, Madden's adaptive procedure apparently drove the listeners to extremely large frequency increments between steps. We assume that they were using a different cue to reach criterion when the increment per step was over 100 Hz.

2 Detection of sinusoidal plus ramp FM

There are some concerns about the influence of subtle spectral differences between the glide and the step signals with the use of the STEP-GLIDE discrimination task. To minimize the possible contamination of "splatter" at each step, Madden used a 17-STEP signal as the standard rather than a true linear glide. Further, the transitions were "rounded" to reduce the spread of energy when the signal frequency was abruptly changed to a new value.

Consider the STEP signal used in the previous work. It can be described as a triangular wave modulator added to a linear ramp before the combined waveform is used to modulate the frequency of a carrier tone. It can be difficult to specify the modulation index of such a combined modulation waveform. If, instead, a sinusoid is added to a linear ramp to produce the combined modulator, the resulting modulator is easy to specify. This new target signal replaces the STEP signal used previously. If the slope of the ramp is zero (i.e., no change in base frequency), the listener's task is simply detection of sinusoidal FM. When the sinusoidal FM is added to a linear ramp, the listener's task is similar to that in the discrimination of STEP versus GLIDE signals (Zhang et al., 1994).

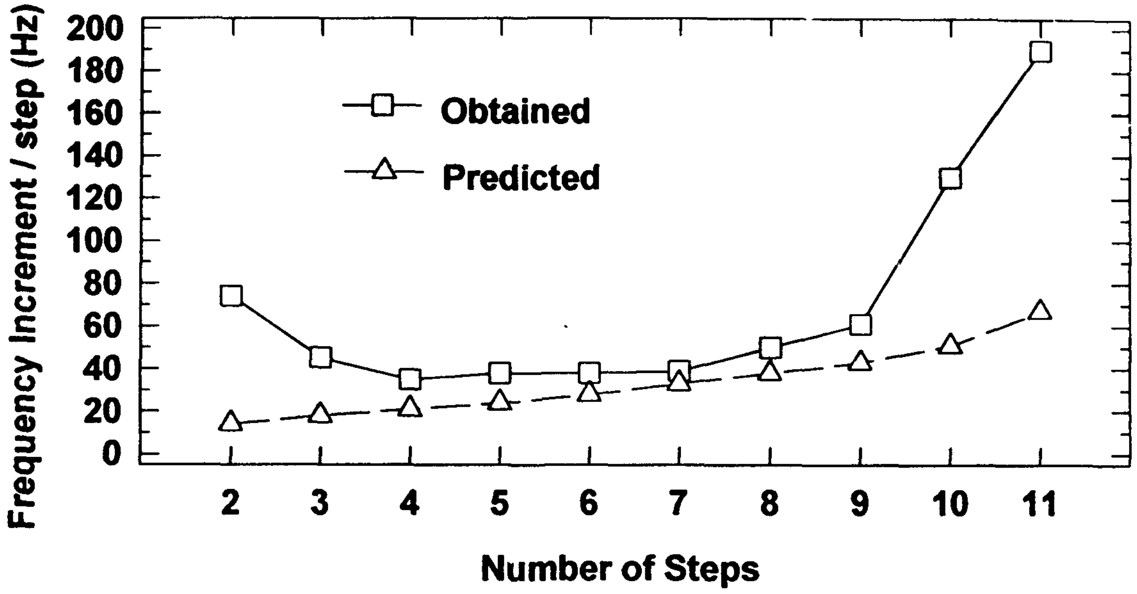


Figure 2. Comparison of Madden's results at 1 kHz with ST-IWAIF model predictions.

2.1 Method

Listeners with normal hearing were asked to determine which of two tones was sinusoidally frequency modulated. In one set of experimental conditions, the standard signal was a steady tone, and the target signal was generated by frequency modulating the standard with a sinusoid at 4, 8, 16, 32, 64, 128 or 256 Hz. In other conditions, the standard signal was frequency modulated by a linear ramp. The target was then generated by adding the sinusoidal FM to the linear FM. To avoid the possibility of anchoring effects, the frequency of each signal was chosen from a uniform random distribution. This is commonly called a roving-frequency condition.

2.2 Signal generation

All signals were generated using the TDT System II. Three listeners were tested at one time. Separate channels from the four-channel D-to-A converter delivered signals to one side of a Sennheiser HD 414 headset. Individual detection thresholds for the standard signals were determined using an adaptive 2AFC procedure; the FM detection task was conducted at 50 dB SL. Signal duration was 250 ms with rise-fall times of 5 ms.

2.3 Procedures

Data were collected in blocks of 50 trials using an adaptive 2Q, 2AFC procedure. A 3-up, 1-down rule was used (Levitt, 1971) to estimate the 79.4 % point on the listeners' psychometric functions. Data were collected from at least six blocks of trials before averaging the results. When the results appeared to be too variable, an additional three blocks were run and the "best" six were averaged.

3 Results

3.1 Sinusoidal FM detection: with or without glide

Figure 3 shows the results for detection of sinusoidal FM at f_c = 1 kHz averaged for three listeners. The abscissa displays modulation frequency, ranging from 4 to 256 Hz. The ordinate is β, the index of modulation required to obtain 79.4% correct detection. Results for detection of sinusoidal FM added to a linear ramp FM are also shown. Here the ramp rises 800 Hz over 250 ms. In general, listeners have more difficulty detecting the sinusoidal FM in the presence of the ramp than they do when the standard is an unmodulated tone. The data displayed in Figure 3 were averaged over the four roving-frequency ranges tested. Remarkably, there is no effect of roving on the listeners' ability to detect the sinusoidal FM, either for the steady base-line, or for the ramped one.

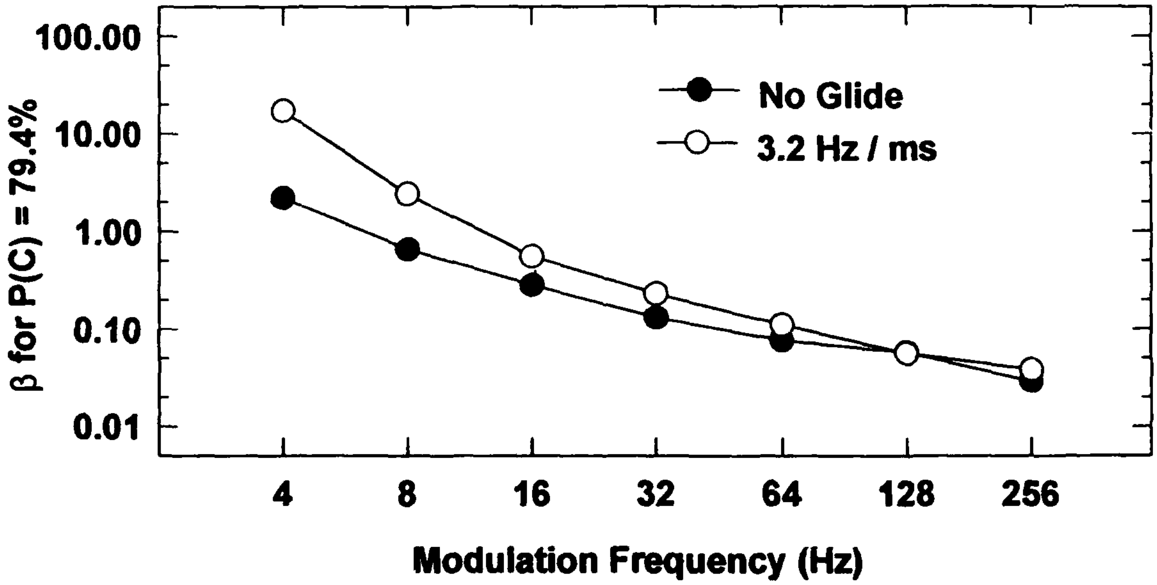


Figure 3. Detection of sinusoidal FM modulation averaged for three listeners. Circles are for simple FM detection; filled squares are for detection of the sine FM added to a linear glide. Starting frequency rove range: none, 200, 400 and 800 Hz.

3.2 Sine wave analog to STEP-GLIDE discrimination

Three listeners with normal hearing were tested in an FM detection task constrained to be analogous to the earlier STEP-GLIDE discrimination task. The STEP signal was replaced by a sinusoidal FM plus GLIDE modulator. Signal duration was set to 100 ms and the transition rate was 4 Hz/ms. The sinusoidal FM was added to the linear ramp with starting phase at 180^0 to better approximate the STEP signal. Detection thresholds for the sine FM were obtained at octave frequencies from 250 Hz through 4 kHz, plus 6 kHz. An adaptive, 3-up, 1-down rule (Levitt, 1971) was used to adjust the period of the sinusoidal FM to approximate the discrete steps used earlier. Thus, for a 100 ms signal, a modulation rate of 10 Hz completes one cycle in 100 ms. To approximate 2 steps, the rate was changed to 20 Hz. For each new modulation rate, the amplitude of the sinusoid was adjusted so that it would have the same power as the triangular modulator that produced the original step function. This is only one of several constraints that might be placed on the sinusoid to "match" it to the triangular waveform.

Discrimination results for three listeners are shown in Figure 4 along with the performance predicted for each one using the ST-IWAIF model. As with the earlier STEP-GLIDE results, the just-detectable period is approximately uniform from 250 Hz to 2 kHz. Performance is poorer at 4 kHz but at 6 kHz it appears to have leveled off.

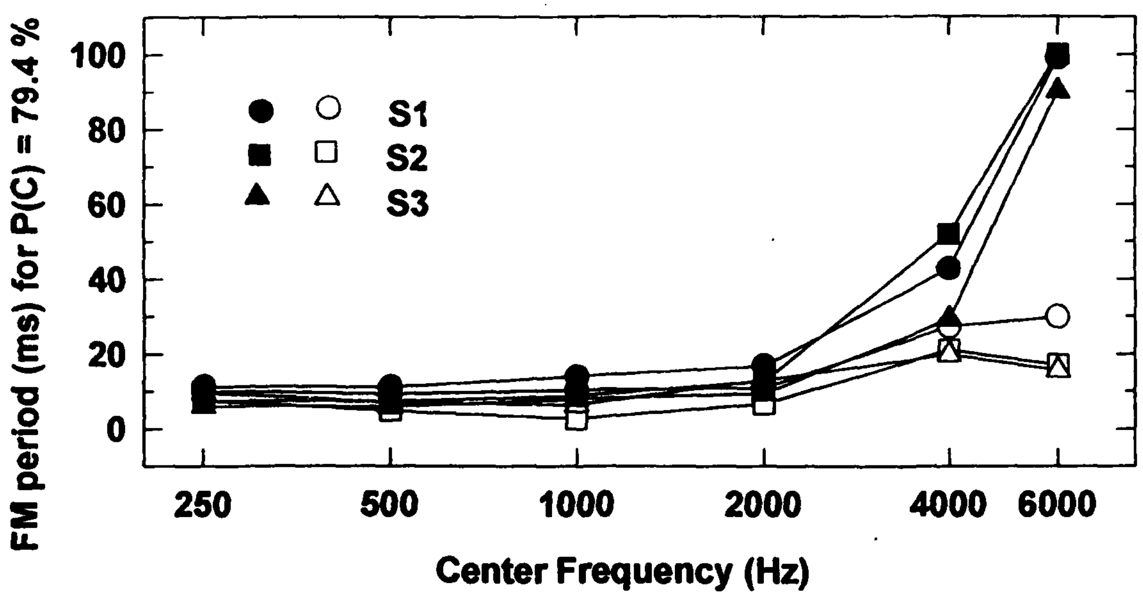


Figure 4. Individual results for the listeners detecting a sinusoidal FM added to a linear glide. Filled symbols are ST-IWAIF predictions for listener performance.

4 Discussion

Our original set of STEP-GLIDE signals led us to conclude that the width of the temporal processing window was approximately constant across signal frequencies below 2 kHz. We attributed listeners' poorer performance in STEP-GLIDE discrimination at higher frequencies to the ear's inability to follow the frequency transition, perhaps due to the loss of phase-locking in the primary neural units. Madden's extension of that work (1994) introduced an adaptive testing procedure and modeled the results using the "traditional" intensity based model used in previous temporal acuity measurements such as gap detection or temporal modulation transfer functions. Madden's adaptive procedure kept the signal duration and number of steps fixed over a block of experimental trials and varied the frequency excursion of the signal pairs. The adaptive procedure determined the frequency interval (FI) for which listeners could distinguish the STEP from the GLIDE. Recent work by Hsu (1993) has shown that the just-discriminable change in frequency-transition slope obeys Weber's law. That is, ΔF/F in Madden's procedure changes with each change required by the adaptive rule.

The intensity model used by Madden matched listener performance fairly well for the mid-range of signal frequencies tested. The model had to be allowed poorer detector efficiency to reproduce the upturn in just-discriminable frequency increment at the smallest and largest number of steps. The large increase in frequency increment required to achieve discrimination performance at 79.4% for the large number of steps is difficult to explain within the limits of the assumed task presented to the listeners. In essence, they are asked to distinguish a transition containing 9, 10 or 11 discrete steps over 50 ms from a standard containing 17 steps over that same duration. The very large frequency increment required to "satisfy" the adaptive procedure suggests that perhaps at this end of the discrimination task, listeners were using a different cue to make the distinction. Certainly, the increments were large enough to suggest that simple frequency discrimination of the initial frequency jump might be the cue.

At the other end of Madden's FI vs. step-number curve, there is a drop in FI with increasing numbers of steps. For 2, 3 and 4 steps the overall frequency excursion is approximately 150 Hz. In the earlier work, we found that the duration of a single step remained constant for criterion performance over a wide range of signal duration and frequency excursions. The falling FI vs. step-number curve may not reflect that same behavior.

We have shown above that a ST-IWAIF model can account for Madden's results about as well as the intensity-based model he proposed. There remain some reservations about signal artifacts (energy splatter at the transitions) and problems with the adaptive procedure as implemented by Madden. Thus, we have proposed that sinusoidal FM added to a linear ramp be used to further test the IWAIF model.

Our initial results indicate that the ST-IWAIF model can account for listener performance in the detection of sinusoidal FM. When the ramp has zero slope, our task is the familiar FMDL task. For that task, we have indicated that introducing a roving-frequency paradigm has little or no effect on listener performance. Our results appear to be reasonable when compared with previous determinations of the FMDL (Moore and Glasberg, 1989) given differences in psychophysical procedures and signal parameters.

When the sinusoidal FM is added to a ramp modulation, listener performance in the detection of FM is somewhat poorer at the lowest FM rates. In the "traditional" FMDL task, the baseline for comparison is fixed in frequency over the duration of the signal. For our ramp-plus-sine FM task, the baseline is moving.

Finally, when we use the sinusoidal FM plus GLIDE signal to replicate the earlier STEP-GLIDE discrimination results, listener performance is consistent with our earlier findings. We suggest that the listener's inability to "follow" the changing baseline at higher frequencies (perhaps due to the loss of phase locking in the auditory nerve) probably accounts for the poorer performance in both STEP-GLIDE discrimination and in FM detection. The ST-IWAIF model predicts progressively poorer performance above 2 kHz than our listeners achieve. This may indicate that such discriminations are not based on the listeners' ability to follow rapid frequency transitions, as the model assumes. However, since the model uses the listener's DLF to estimate variance in the tracking task, model performance is degraded as the DLF increases at higher signal frequencies.

5 Acknowledgment

This work was supported by a grant from the AFOSR.

6 References

Anantharaman, J. N., Krishnamurthy, A. K. and Feth, L. L. (1993) Intensity weighted average of instantaneous frequency as a model for frequency discrimination. J. Acoust. Soc. Amer. 94, 723-729.

Chistovitch, L. A. and Lubinskaya, V. V. (1979) The 'center of gravity' effect in vowel spectra and critical distance between formants: Psychoacoustical study of perception of vowel-like stimuli. Hearing Research 1, 185-195.

Cullen, J. K., Houtsma, A. J. M. and Collier, R. (1992) Discrimination of brief tone-glides with high rates of frequency change. Fifteenth Midwinter Research Meeting of the ARO, 67.

Feth, L. L. (1974) Frequency discrimination of complex periodic tones. Perception and Psychophysics 15, 375-378.

Feth, L. L., Neill, M. E., and Krishnamurthy, A. K. (1989) Auditory temporal acuity for dynamic signals. J. Acoust. Soc. Amer. 86, S122-123.

Feth, L. L., O'Malley, H. and Ramsey, Jr. J. (1982) Pitch of unresolved, two component complex tones. J. Acoust. Soc. Amer. 72, 1403-1412.

Fineberg, A. B. Mammone, R. J., and Flannagan, J. L. (1992) Application of the modulation model to speech recognition. In Proceedings of the ICASSP, I-541-I544.

Forrest, T. G. and Green, D. M. (1987) Detection of partially filled gaps in noise and the temporal modulation transfer function. J. Acoust. Soc. Amer. 82, 1933-1943.

Green, D. M. (1971) Temporal auditory acuity. Psychological Review 78, 540-551.

Green, D. M. (1973a) Minimum integration time. In Basic Mechanisms in Hearing. H. Duifuis, J. W. Horst and H. P. Wit (eds.), 829-846, Academic Press NY.

Green, D. M. (1973b) Temporal acuity as a function of frequency. J. Acoust. Soc. Amer. 54, 373-379.

Green, D. M. (1985) Temporal factors in psychoacoustics. In Time Resolution in Auditory Systems A Michelsen (ed.), 122-140, Springer Verlag NY.

Green, D. M. (1985) Profile Analysis: Auditory Intensity Discrimination. Oxford Science Publications NY.

Green, D. M. and Forrest, T. G. (1988) Detection of amplitude modulation and gaps in noise. In Basic Mechanisms in Hearing. H. Duifuis, J. W. Horst and H. P. Wit (eds.), 323-331, Academic Press NY.

Hsu, C-Y. (1993) Auditory discrimination of frequency transitions by human listeners and a computational model. Unpublished Ph. D. Dissertation, Ohio State University.

Krishnamurthy, A. K. and Feth, L. L. (1993) Short-term IWAIF model for frequency discrimination. J. Acoust. Soc. Amer. 93, 2387.

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Amer. 49, 476-477.

Liberman, A. M., Delattre, P. C., Gerstman, L. J. and Cooper, F. S. (1956) Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exptl. Psych. 52, 127-137.

Madden, J. P. (1994) The role of frequency resolution and temporal resolution in the detection of frequency modulation. J. Acoust. Soc. Amer. 95, 454-462.

Madden, J. P. and Feth, L. L. (1992) Temporal resolution in normal-hearing and hearing-impaired listeners using frequency-modulated stimuli. J. Speech Hear. Res. 35, 436-442.

Maragos, P. Kaiser, J. F. and Quatieri, T. F. (1992) On separating amplitude from frequency modulations using energy operators. In Proceedings of the ICASSP, II-1 - II-4.

Moore, B. C. J., Glasberg, B. R., Plack, C. J. and Biswas, A. K. (1988) The shape of the ear's temporal window. J. Acoust. Soc. Amer. 83, 1102-1116.

Moore, B. C. J. and Glasberg, B. R. (1989) Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. J. Acoust. Soc. Amer. 86, 1722-1732.

Plack, C. J. and Moore, B. C. J. (1990) Temporal window shape as a function of frequency and level. J. Acoust. Soc. Amer. 87, 2178-2187.

Plack, C. J. and Moore, B. C. J. (1991) Decrement detection in normal and impaired ears. J. Acoust. Soc. Amer. 90, 3069-3076.

Teager, H. M. and Teager, S. M. (1990) Evidence for non-linear sound production mechanisms in the larynx. In NATO Advanced Study Institute on Speech Production and Speech Modeling, 241-261, Kluwer Press.

Viemeister, N. F. (1979) Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Amer. 66, 1354-1380.

Zhang, T., Feth, L. L. and Krishnamurthy, A. K. (1994) Detection of frequency modulation in steady and gliding tones. J. Acoust. Soc. Amer. 95, 2965.

FIRST INTERNATIONAL CONFERENCE ON ELECTRONICS, CIRCUITS AND SYSTEMS, DEC 19-21, CAIRO, EGYPT.

A Multi-channel Intensity-Weighted Average of Instantaneous Frequency Model

Mohamed A. Mokheimer[1], Jayanath N. Anantharaman[1],
Ashok K. Krishnamurthy[1] and Lawrence L. Feth[2]

[1]Department of Electrical Engineering
[2]Department of Speech and Hearing Science
The Ohio State University
2015 Neil Avenue
Columbus, Ohio 43210

Abstract

The Intensity Weighted Average of Instantaneous Frequency (IWAIF) model has been successfully used to explain a number of psychoacoustic results in which the primary cue used by the listener is frequency. The IWAIF of a signal is the frequency of the center-of-gravity of the positive frequency half of the signal spectrum. With a few exceptions, the IWAIF model has been applied only to narrowband signals. In this paper, we propose a multi-channel extension of the IWAIF model that is useful in analyzing wideband signals. The output of the multi-channel IWAIF model is a vector of IWAIF (frequency) values. We then present two applications of the IWAIF/multi-channel IWAIF model: (1) to explain the Chistovich "perceptual formant" effect observed in vowel perception; and, (2) to model the detection of mixed amplitude and frequency modulation (MM) by human listeners. Comparisons of the predictions of the model with psychoacoustic data show that the model predictions are in reasonable agreement with the data at high modulation rates (256 Hz), while at lower modulation rates (4 Hz, 16 Hz), the model predicts a phase dependence that is not present in the data. We speculate that at low modulation frequencies, a short-term IWAIF model may be more appropriate.

Introduction

Our work is based on the underlying assumption that human auditory communication can be modeled as a modulation–demodulation process. In this model, the information in the sound pressure wave is encoded as variations in the amplitude and frequency of the signal. Speech, music and most environmentally important sounds can be described in this way. The human listener's task, then, is to demodulate the sound stream to extract the encoded information. This modulation–demodulation view of auditory processing has also been recently (and independently) advocated by Maragos et al. (1992), who have proposed a non-linear energy operator for extracting modulation information from signals. Fineberg et al. (1992) have also applied a modulation model to speech recognition.

Our work in the past several years has concentrated primarily on the processing of narrowband frequency modulated signals. Feth (1974) proposed the Envelope Weighted Average of Instantaneous Frequency (EWAIF) as a model for the processing of such FM as well as AM signals. Recently, we have developed the Intensity Weighted Average of Instantaneous Frequency (IWAIF) as an alternative to the EWAIF (Anantharaman et al., 1993). The IWAIF has a number of advantages over the EWAIF: (i) it is easier to compute; and (ii) it has an intuitive frequency domain interpretation, since the IWAIF of a signal is simply the center-of-gravity frequency of the signal spectrum (Anantharaman, 1992).

The primary applications of the EWAIF and IWAIF models have been to narrowband signals that are confined to a single critical band. Most real-world signals such as speech and music are wideband. The relevant information in such signals, such as harmonicity, correlated amplitude or frequency modulation etc., is often spread over several critical bands. The psychoacoustic phenomena of Comodulation Masking Release and Profile Analysis illustrate that the human auditory system is capable of following a multi-component signal in several channels simultaneously (Green, 1988). It is essential to extend the IWAIF model to a multi-channel version to analyze such wideband signals. Any extension of the IWAIF model to wideband signals will have to incorporate our existing knowledge of the auditory system by including such features as the basilar membrane filtering, compression, short- and long-term adaptation, phase locking etc. As a first step in this direction, we present in this paper a multi-channel

IWAIF model that includes basilar membrane filtering and spatial integration. This model is applied to two psychoacoustic results: (i) the "perceptual formant" effect in vowel perception described by Chistovich (1979, 1985); and (ii) mixed modulation (MM) perception.

The next section describes the basic IWAIF model; then we introduce the multi-channel IWAIF model. The applications of the multi-channel IWAIF model to vowel perception and modulation detection form the next two sections, and we conclude with some recommendations for future work.

Basic IWAIF Model

Let $s(t)$ be a real signal, with instantaneous envelope $e(t)$ and instantaneous frequency $f(t)$. The IWAIF of the signal $s(t)$ is defined as (Anantharaman et al. 1992),

$$\mathrm{IWAIF}[s(t)] = \frac{\int_0^T e^2(t)\, f(t)\, dt}{\int_0^T e^2(t)\, dt} \qquad (1)$$

A much more convenient representation of the IWAIF of $s(t)$ is obtained if the above expression is transformed to the frequency domain. As shown by Anantharaman et al. (1993),

$$\mathrm{IWAIF}[s(t)] = \frac{\int_0^\infty f\, |S(f)|^2\, df}{\int_0^\infty |S(f)|^2\, df} \qquad (2)$$

where $S(f)$ is the Fourier transform of $s(t)$. Thus, the IWAIF of a real signal is the "center of gravity" frequency of the positive portion of its energy density spectrum. The IWAIF can be computed very efficiently using the FFT (Anantharaman et al. 1993).

Multi-channel IWAIF Model

The multi-channel IWAIF model consists of three stages (Mokheimer 1993):

1.A filterbank stage, that models the bandpass filtering of the basilar membrane on the incoming signal. We use the Gammatone filterbank proposed by Patterson et al. (1987) for this purpose.

2. A Spatial integration stage, that combines the output of a number of adjacent filters. As presently configured, the output from three adjacent filters are combined. The decision to combine only three adjacent output was based on the observation that the "perceptual formant" effect in vowel perception only occurs if two formants are less than 3 critical bands apart (Chistovich, 1979, 1985).

3. An IWAIF/Intensity computation stage, that computes the IWAIF and the intensity at the output of each channel.

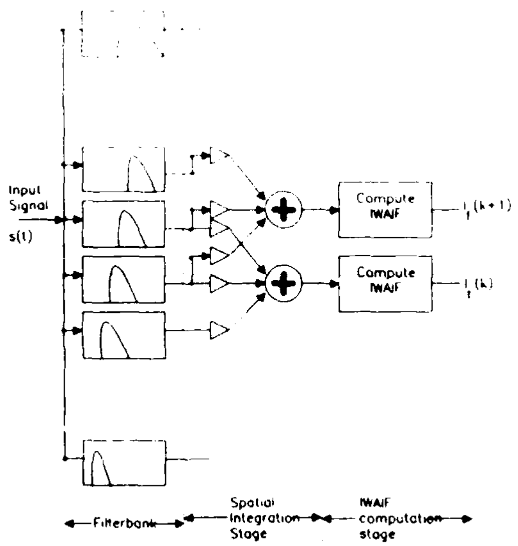


Figure 1: The proposed multi-channel IWAIF model.

Figure 1 shows the proposed Multi-channel IWAIF model. The blending weight for each channel in combining the output of adjacent channels is chosen to be the relative intensity of that channel. Experiments with various weighting choices showed that these lead to the best results for the modulation detection task.

Given the signal $s(t)$, $0 \leq t \leq T$, the model leads to a vector of (IWAIF, Intensity) pairs, one pair for each channel; i.e., $(I[n], L[n])$, $n = 1, ..., N_f$, where N_f is the number of frequency channels, $I[n]$ is the IWAIF value for the n^{th} channel, and $L[n]$ is the intensity level for the n^{th} channel.

Fig. 2 shows the output of the multi-channel IWAIF model to a sine wave at 1000 Hz. Notice that an IWAIF value is computed even for channels with center frequencies far from 1000 Hz, whose output intensity is very small. This is because the IWAIF itself is independent of signal energy, and the IWAIF value computed in these channels is dominated by round-off noise and the filter impulse response. We make the assumption that the auditory system, in most situations, ignores channels with relatively low energy. Thus, we only retain for further processing those channels whose relative intensity is within 35 dB of the maximum intensity. Notice that for these channels, the computed IWAIF is very close to 1000 Hz.

Modeling the "Perceptual Formant" effect using the IWAIF

In a series of papers on the perception of vowel quality, Chistovich and her colleagues (1979, 1985) asked

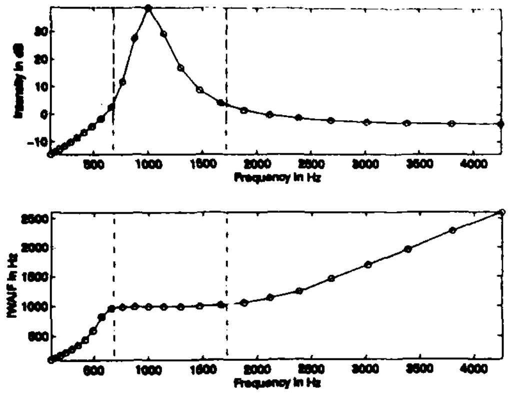


Figure 2: Output of the multi-channel IWAIF model to a tone at 1000 Hz. The intensity at the output of a channel is plotted against the channel center frequency in the top graph; the bottom graph shows the IWAIF value against the center frequency. The vertical dashed lines bracket the set of channels in which the intensity is within 35 dB of the maximum.

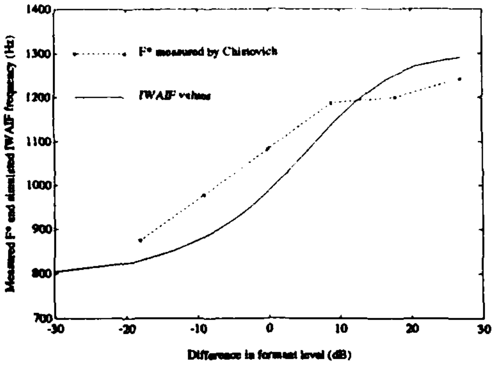


Figure 3: This figure compares the IWAIF frequency of synthetic two formant stimuli (solid line) with the center frequency of a synthetic one-formant stimulus that Chistovich's listeners matched to the two formant stimulus. The abscissa is the difference in level between th. formants in the two formant stimuli.

listeners to match synthesized two formants stimuli to a single formant stimulus whose center frequency could be changed. The levels of the two formants in the two formant stimulus were varied. Chistovich and her colleagues found that as long as the two formants are less than 3.5 critical bands apart, listener's matched the two formant stimulus to a single formant stimulus whose center frequency was equal to the frequency of the center of gravity of the two formant stimulus. Subsequent work by others (Beddor and Hawkins, 1990) has lead to a better understanding of the factors that govern the "perceptual formant" effect, but the existence of the effect itself is generally accepted. Chistovich suggests that the auditory system performs a spatial (i.e. spectral) integration over wide intervals of the cochlea, which leads to the "perceptual formant" effect. Using steady-state vowels sounds, she demonstrated that this model predicted listener performance in vowel recognition experiments very well. As explained earlier, the IWAIF of a signal is the "center of gravity" of the positive frequency half of the spectrum; hence the "perceptual formant" should be at the IWAIF frequency. In Figure 3 we compare the results of Chistovich for the frequency of the "perceptual formant" that listeners matched for two formant vowels, with the IWAIF. As can be seen, the agreement is quite good.

Application of the Multichannel IWAIF Model to Mixed Modulation Detection

The human auditory system appears to use modulation as an important cue in grouping the separate components of a signal. We can easily detect both amplitude and frequency modulation, and combinations of both, called mixed modulation (MM). Psychoacoustic studies suggest that fairly complex perception mechanisms, which depend on the type and frequency of modulation (Ozimek and Sek, 1987; Moore and Sek, 1992), are involved in detecting MM signals. A number of models for modulation detection have been proposed (Hartman, 1982, Moore, 1989, Zwicker 1962 and Florentine 1981), but each is unable to account for some of the psychoacoustic data.

Following Ozimek and Sek (1987) and Hartmann and Hnath (1982), the MM signal can be written as:

$$a(t) = A_o(1 + m\cos\omega_m t)\sin(\omega_o t + \beta\sin(\omega_m t + \phi)), \tag{3}$$

where
A_o=carrier amplitude, ω_m=modulation angular frequency, ω_o=carrier angular frequency, $\Delta\omega$=maximum frequency deviation, m = AM modulation index, $\beta = \Delta\omega/\omega_m$=FM modulation index, and ϕ = relative phase angle between AM and FM.

Assuming that $m\beta \ll 1$, the spectrum of a MM signal consists of three components, the central of which represents the carrier, while the sidebands are due to the modulation effect. The amplitudes and phases of the sidebands depend on the phase shift between the signals that modulate the amplitude and frequency of the carrier, and the relative levels of AM and FM..

The insets in Figure 4 show the waveforms (top) and schematics of the spectral magnitude (bottom) of mixed modulation signals with different relative phase angles ϕ between AM and FM. Other parameters of these signals are: carrier frequency = 1000 Hz, modulating frequency = 256 Hz, m = 0.1 and $\beta = 0.1$. Also shown in Fig. 4 are the multi-channel IWAIF values for the carrier alone (filled circles) and the MM signal (open squares). Only those channels whose intensity is within 35 dB of the maximum are shown in the figure.

Fig. 4 clearly shows how the phase angle ϕ effects the degree of asymmetry of the MM spectra, which is in turn reflected in the multi-channel IWAIF. As an example for case $\phi = 0$ the amplitude of the upper sideband is higher than the amplitude of the lower sideband. Consequently, the IWAIF values at the output of the channels centered at frequencies higher than the carrier frequency (high frequency side channels) are perturbed further from the carrier frequency than the IWAIF values for the low frequency side channels. A similar explanation applies to the other cases of ϕ.

To derive quantitative results comparing the predictions of the multi-channel IWAIF model to listener performance in detecting MM signals, it is necessary to choose an apropriate detection model. We have adapted the multi-channel detector proposed by Durlach et al (1986) for this purpose. The detector model is based on the following assumptions: (i) The signal is detected by the changes in the IWAIF values it produces in different channels; and (ii) Internal noise is present in each channel, and is added after IWAIF computation i.e. it serves to perturb the computed IWAIF values. The noise is assumed to be zero-mean, Gaussian, and statistically independent across channels, and to be independent of the stimulus. The variance of the noise is assumed to be frequency-dependent.

Under these assumptions, following Durlach et al., the sensitivity d' is given by

$$d' = K[\sum_{i=1}^{Nm} \frac{[I_1(i) - I_2(i)]^2}{\sigma_F^2(i)}]^{\frac{1}{2}} \qquad (4)$$

where I_1(i) and I_2(i) are the IWAIF values in the ith frequency channel for the signals to be discriminated, and $\sigma_F(i)$ is the noise standard deviation for the ith frquency channel, N_m represents the number of the channels of interest with sufficiently high intensity, and K is free parameter that represents the efficiency of the detection mechanism. Since the IWAIF is basically a frequency value, σ_F(i) can be chosen the frequency difference limen (FDL) at the center frequency of the ith channel and duration of interest. We used the FDL data available in (Moore, 1974) for different center frequencies and a tone duration of 100 ms to obtain the values of $\sigma_F(i)$. Finally, the parameter K is estimated by matching the prediction of the model to published FM detection data in (Moore and Sek, 1992).

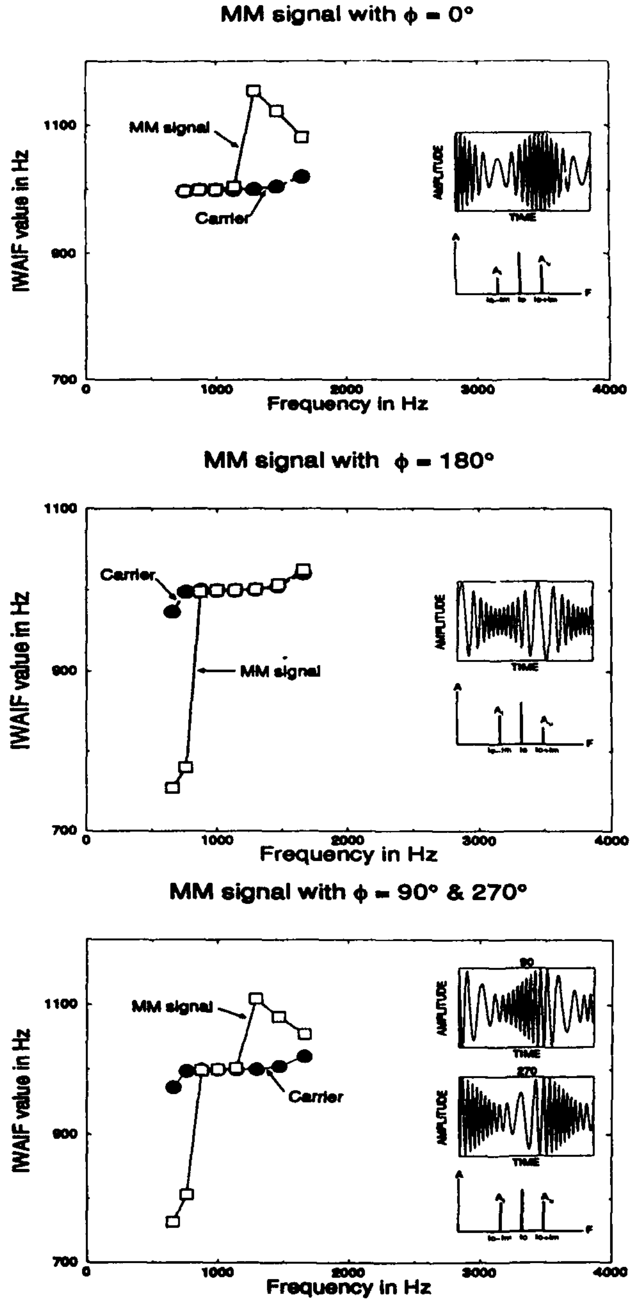


Figure 4: The waveform (top inset), spectral magnitude (bottom inset), and multi-channel IWAIF of MM signal (open squares) with different relative phase angles (ϕ) between AM and FM. Also shown are the multi-channel IWAIF values of the carrier alone (filled circles).

Moore and Sek (1992) investigated the amount of FM (at a fixed phase angle ϕ) that needs to be added to a signal containing a fixed, "sub-threshold" amount of AM, until the resulting MM signal is just discriminable from the carrier signal alone. Listener performance on this task changes considerably depending on the modulating frequency. At low modulation frequencies (4 Hz–16 Hz), the amount of FM needed for detection is nearly independent of the phase ϕ. For a modulation frequency of 256 Hz, on the other hand, the amount of FM needed for detection shows strong phase effects.

Figure 5 compares listener performance (open symbols) with the model prediction (filled symbols) for an FM frequency of 256 Hz and 4 Hz. The data show that at 256 Hz frequency, there were very large effects of the relative phase ϕ. For $\phi = 0$ (circle symbol) , i.e., when the maxima in amplitude and frequency were coincident, the coexisting AM made the FM harder to discriminate from the carrier, i.e., an increase in AM depth (m) caused an increase in the FM index (β) required for threshold. An opposite effect was observed for $\phi = 180$, when maxima in amplitude and minima of the frequency were coincident: an increase in m caused a significant decrease of β required for threshold. For $\phi = 90$ or 270, the value of β required for threshold, decrease slightly with increasing m. This supports the idea that the spectral structure of the modulated signal and the frequency selectivity of the auditory system are the bases for the discrimination, and the temporal fine structure (i.e., changes in frequency and amplitude over time) does not play any role in the detection process. The multichannel IWAIF predictions in this case agree quite well with the data.

The results at the lowest modulation frequency (4 Hz) are also shown in Fig. 5. Here the listener data (open circles) show no clear effect of the relative phase ϕ. This result was tested with the multichannel IWAIF model. The model predicts an effect of the relative phase that was not observed in the data; at the same time, for $\phi = 90$ and 270, the model predicts that a lesser amount of FM modulation is needed as compared to the listener data.

Discussion

The multi-channel IWAIF model predicts listener detection of MM signals at high modulation rates quite well, but fails at the lower modulation frequencies. This is also true of other models, such as the one proposed by Hartmann and Hnath (1982). A common feature of both these models is that they use only the spectral properties of the signal and ignore the temporal structure. We speculate that at very low

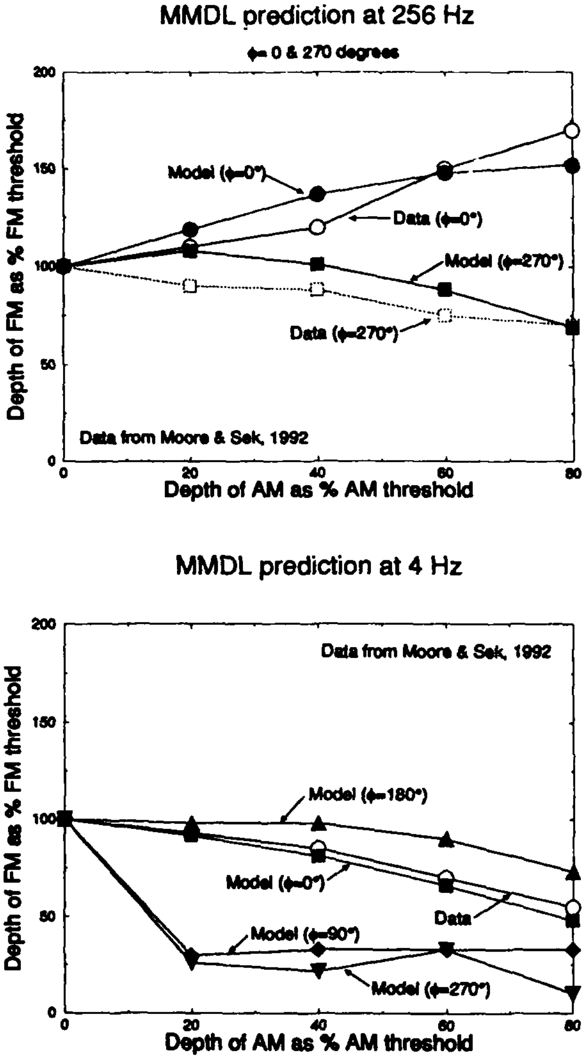


Figure 5: Comparison of listener performance with multi-channel IWAIF model predictions for the discrmination of a MM signal from the carrier. The modulation frequency is 256 Hz for the top graph and 4 Hz for the bottom graph.

modulation rates (4 Hz), when listeners can easily follow the temporal structure of the signal, detection is determined by temporal properties. At moderate modulation rates (16 Hz–64 Hz), perhaps a combination of temporal and spectral effects contribute to detection. We believe that a short-term IWAIF model may be more applicable at the low modulation rates (Krishnamurthy and Feth, 1993).

Conclusions

We have presented a multi-channel IWAIF model that is applicable to wideband signals, and incorporates basilar membrane filtering and spatial integration. An application of the model to the detection of mixed modulation was described. The results indicate thatthe model matches listener performance at high modulation rates. We plan to extend the model to include more stages of auditory processing such as temporal integration. Also, we will combine the multi-channel IWAIF and the short-term IWAIF models leading to a model that is useful for time-varying, broadband signals.

Acknowledgements

This work was supported in part by a grant from the Air Force Office of Scientific Research.

References

[1] Anantharaman, J. N. (1992). "A multichannel signal processing model for comlex sound discrimination," Master'thesis, The Ohio State University,

[2] Anantharaman, J. N., Krishnamurthy, A. K. and Feth, L. L (1993). "Intensity-weighted average of instantaneous frequency as a model for frequency discrimination," Journal of the Acoustical Society of America, vol.94, pp. 723-729.

[3] Beddor, P. S. and Hawkins, S. (1990). "The influence of spectral prominence on perceived vowel quality". Journal of the Acoustical Society of America, 87(6): pp2684-2704.

[4] Chistovitch, L. A. and Lublinskaya, V. V. (1979) "The 'center of gravity' effect in vowel spectra abd critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hearing Research, vol.1, pp.185-195.

[5] Chistovitch, L. A. (1985). Central auditory processing of peripheral vowel spectra. Journal of the Acoustical Society of America, 77: pp798-805.

[6] Durlach, N. L., Braida, L. D. and Ito, Y. (1986). "Towards a model for discrimination of broadband signals". Journal of the Acoustical Society of America, vol.80, pp. 63-70.

[7] Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," Perception and Psychophysics, 15(2), pp 375-378.

[8] Fineberg, A. B., Mammone, R. J., and Flanagan, J. L. (1992). "Application of the modulation model to speech recognition". Proceedings of the international Conference on Acoustics, Speech, and Signal Processing, pp.I-541-I544, San Francisco, CA.

[9] Florentine, M. (1981). "An excitation-pattern model for intensity discrimination". Journal of the Acoustical Society of America, vol.70, pp. 1546-1654.

[10] Green, D. M. (1988). "Profile Analysis: Auditory Intensity Discrimination". Oxford Science Publications, New York.

[11] Hartmann, W. M. and Hnath, G. M. (1982). "Detection of Mixed Modulation". Acoustica, vol. 50, no. 5, pp. 297-312.

[12] Krishnamurthy, A. K. and Feth, L. L. (1993). "Short-term IWAIF model for frequency discrimination," presented in the 125th Meeting of ASA, Ottawa-Canada.

[13] Maragos, P., Kaiser, J. F., and Quatieri, T. F.(1992) "On separating amplitude from frequency modulations using energy operators," Proceedings of the International conference on Acoustics, Speech, and Signal Processing, pp II-1–II-4, San Francisco, CA

[14] Mokheimer, M. A. (1993) "Multichannel Iwaif Model for Modulated Signals," Master'thesis, The Ohio State University.

[15] Moore, B. C. J. (1974). "Relation between the critical bandwidth and the frequency difference limen". Journal of the Acoustical Society of America, vol.55, pp.359.

[16] Moore, B. C. J. and Glasberg, B. R. (1989). "Mechanism underlying the frequency discrimination of pulsed tones and the dtection of frequency modulation". Journal of the Acoustical Society of America, vol.86, pp.1722-1731.

[17] Moore, B. C. J and Sek, A. (1992). "Detection of combined frequency and amplitude modulation". Journal of the Acoustical Society of America, vol.92, pp.3119-3131.

[18] Ozimek, E. and Sek, A. (1987). "Perception of amplitude and frequency modulated signals (mixed modulation)". Journal of the Acoustical Society of America, vol.82, pp.1598-1603.

[19] Patterson, R. D., Nimmo, S.I, Holdworth, J. and Rice, P. (1987). "An efficient auditory filterbank based on the gammatone function". Annex B of the SVOS Final report (Part A: The auditory Filter Bank).

[20] Wier, C. C, Jesteadt, W. and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level". Journal of the Acoustical Society of America, vol.61, pp. 178-184.

[21] Zwicker, E. (1962). "Direct comparisons between the sensations produced by frequency modulation and amplitude modulation". Journal of the Acoustical Society of America, vol.34, pp.1425-1430.